DOCUMENT RESUME

ED 369 814 TM 021 452

AUTHOR Koeller, Olaf

TITLE Identification of Guessing Behavior on the Basis of the Mixed Rasch Model.

PUB DATE Apr 94

NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Achievement Tests; *Cognitive Ability; Difficulty Level; Grade 7; *Guessing (Tests); Item Response Theory; Junior High Schools; *Junior High School Students; *Multiple Choice Tests; Problem Solving; Simulation; Test Items; Test Validity; True Scores

IDENTIFIERS Empirical Research; Latent Class Analysis; *Rasch Model

ABSTRACT

Scholastic achievement tests and mental ability tests normally consist of a set of multiple choice items, all of which are assumed to measure school-relevant cognitive abilities. The presumption, in a given test situation, is that the answers/solutions to the given tasks represent cognitive capabilities on the part of the examinees. The purpose of this paper is to show that this assumption does not always hold. Analyzing simulated and empirical data it is proved that, based on the mixed Rasch model (Rost, 1990), it is possible to identify those examinees who have applied a guessing strategy to solve multiple choice items. As an empirical example the results in a biology test consisting of 23 items, each having 5 choices were analyzed. On the basis of the responses from 5,641 7th grade students a guessing class was identified. Further analyses provided information indicating that guessing behavior is shown by students with lower-level cognitive abilities, who might have used the "random strategy" to cope with the items that were too difficult. Five tables and four figures provide study data. (Contains 21 references.) (Author/AA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 369 814

Identification of Guessing Behavior on the Basis of the Mixed Rasch Model

Olaf Koeller[1]

Institute for Science Education (IPN) at the University of Kiel

Scholastic achievement tests and mental ability tests normally consist of a set of multiple choice items, all of which are assumed to measure school-relevant cognitive abilities. The presumption, in a given test situation, is that the answers/solutions to the given tasks represent cognitive capabilities on the part of the examinees. Our current purpose is to show that this assumption does not always hold. Analyzing simulated and empirical data we proved that, based on the mixed Rasch model (Rost, 1990), it is possible to identify those examinees who have applied a guessing strategy to solve multiple choice items. As an empirical example we analyzed the results in a biology test consisting of 23 items, each having 5 choices; and on the basis of the responses from 5641 7th grade students we identified a guessing class. Further analyses provided information indicating that guessing behavior is shown by students with lower-level cognitive abilities, who might have used the "random strategy" to cope with the items that were too difficult.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

OLAF KOELLER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

[1] **Author's address:** Institute for Science Education (IPN) at the University of Kiel, Olshausenstr. 62, D-24098 Kiel, Germany

TM021452

BEST COPY AVAILABLE

1. Introduction

When creating and applying new achievement tests with multiple choice items or questionnaires with rating scales educational researchers assume that these instruments are suitable to measure the same trait for all subjects tested. Researchers are pleased, if ensuing item analyses reveal that a measurement model with only one latent dimension holds for the data, since this facilitates interpretation of individual test scores. Often some examinees of a given sample do not show the expected response behavior, however, for instance in an achievement test where some subjects guess, or in a questionnaire where some subjects show special response sets such as *the tendency toward the mean* or *the tendency toward extreme judgements*. Rost (1994) labled these examinees as "unscalables" due to their deviant response behavior, which is beyond the scope of item response models like the ordinary Rasch model (Rasch, 1960). Rost (1994) Rost and Georg (1991) and Rost and Davier (1993) suggested the mixed Rasch model (MRM) as a powerful tool in dealing with "unscalables" in data sets. The purpose of our current work is to introduce this model as a method for identifying guessing behavior in achievement tests.

Scholastic achievement tests and mental ability tests normally consist of a set of multiple choice items, all of which are assumed to measure school-relevant cognitive abilities. The presumption, in a given test situation, is that the answers/solutions to the given tasks represent cognitive capabilities on the part of the examinees. Koeller, Rost and Koeller (1994) demonstrate that this assumption does not always hold. In their study which dealt with individual differences in solving spatial tasks, the authors administered cube tasks to 2558 7th grade students and, applying a latent class analysis to the data, concluded that some examinees (16% of the whole sample) employed a guessing strategy to solve the tasks.

Based upon these results, our current questions are: Do we find such an undesired guessing behavior in other achievement tests, and is the MRM a suitable statistical method for identifying subjects who guessed? To answer these questions we will first introduce the MRM and other procedures to model guessing behavior in Item Response Theory. Next, we will analyze a simulated and an empirical data set on the basis of the MRM with the PC-program MIRA (Rost & Davier, 1992). Contrary to usual literature, which deals only with the psychometric or statistical issues of guessing, we will place additional emphasis on the relationships between guessing behavior, motivational and cognitive variables. In regard to cognitive variables we will test the

plausible hypothesis that guessing behavior is applied by students with lower-level cognitive abilities.

In the sequel, we will give a short introduction to the MRM. To promote a better understanding of the MRM concept, some basics of the ordinary dichotomous Rasch model (RM; Rasch, 1960) and of the latent class analysis (LCA; Lazarsfeld, 1950; Lazarsfeld & Henry, 1968; Rost, 1988) are first presented.

1.1 The Dichotomous Rasch Model (RM)

Let $p(x_{vi})$ denote the response probability (probability of success) of person v on item i. The main idea of the dichotomous RM is to decompose $p(x_{vi})$ into a linear combination of an item parameter (difficulty) and an individual parameter (person's ability). Since the manifest variable varies only between zero and one, not the probability itself but the logit of this probability is decomposed:

$$\ln\left(\frac{p(x_{vi})}{1-p(x_{vi})}\right) = \xi_v + \alpha_i \qquad (1)$$

i.e.: The logit of the response probability is equal to the sum of a person's ability ξ_v and the item difficulty α_i.

As an easy transformation of Equation (1) the better-known response function of the Rasch model results:

$$p(x_{vi}) = \frac{e^{(\xi_v+\alpha_i)}}{1+e^{(\xi_v+\alpha_i)}} \qquad (2)$$

This relationship between latent variables and the response probability is often represented by the so-called Item Characteristic Curve (ICC), shown in Figure 1.

Although this function is nonlinear, it is evident from Figure 1 that the relationship between a person's ability and the probability of success is nearly linear in most segments of the latent continuum. Procedures for parameter estimation and goodness-of-fit tests are described in the relevant literature (e.g. Hambleton & Swaminathan, 1989; Wright & Masters, 1982). Fundamental assumptions of the RM

are (1) local independence of the items, (2) homogeneity of persons and items and (3) specific objectivity. This last means that item parameter estimations are independent of the group of examinees drawn from the population of examinees and that persons' ability estimations are independent of the particular choice of test items drawn from the population of items. These strong model assumptions are often violated, making other models of Item Response Theory (IRT) more attractive.

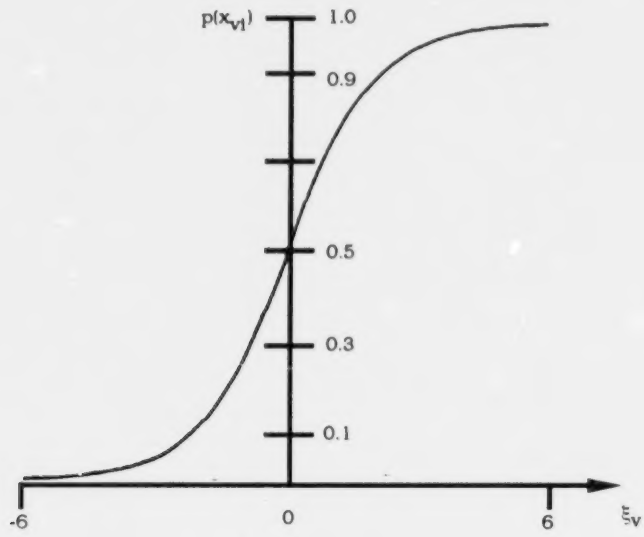

Figure 1
Item Characteristic Curve in the Rasch model

1.2 The Latent Class Analysis (LCA) for Dichotomous Data

The LCA is also an IRT-model, in which the multivariate relations among observed categorical variables are explained by the influence of a latent nominal variable. The response probability of person v on a dichotomous item i is now defined as

$$p(x_{vi}) = \sum_{g}^{G} \pi_g \pi_{i|g}, \tag{3}$$

with restriction

$$\sum_{g}^{G} \pi_g = 1, \tag{4}$$

where G is the number of subpopulations or latent classes, π_g a probability parameter defining the size of class g and $\pi_{i|g}$ the probability of success on item i within class g. In the case of m observed variables the probability of a person's response pattern can be described as

$$p(\mathbf{x}_v) = \sum_{g}^{G} \pi_g \prod_{i}^{m} \pi_{i|g}. \qquad (5)$$

The multiplication of the conditional probabilities $\pi_{i|g}$ within the latent classes follows from the assumption of local independence of the items. The estimation of the unknown parameters π_g and $\pi_{i|g}$ is usually performed with the EM-Algorithm, described in detail by Rost (1988). A model with G classes fits the empirical data perfectly when all observed variables are independent within each class. To assess the model fit of a given solution with G latent classes, two goodness-of-fit statistics are computed: Akaike's Information Criterion and the Best Information Criterion (BIC; Bozdogan, 1987):

$$AIC = -2\ log(L) + 2\ k,$$
$$BIC = -2\ log(L) + log(N)\ k,$$

where L is the maximum of the likelihood-function, k the number of estimated independent parameters and N the sample size; the smaller the goodness-of-fit indices, the better a model with G classes fits. Particularly with larger sample sizes, the BIC appears to produce more valid results than the AIC.

The usual significance tests, i.e. the Pearson χ^2-test for comparing observed and expected pattern frequencies or the likelihood ratio test for comparing different models, are normally not applicable. Both tests have similar asymptotic requirements. The χ^2-test requires expected frequencies greater or equal to one for all possible response patterns, which is usually not fullfilled. In the case of more than 8 dichotomous items 2^8=256 possible response patterns can occur and researchers need very large sample sizes to apply the χ^2-test to the data. Unfortunately, the same problem arises when applying a likelihood-ratio test to compare two different class solutions. The likelihood-ratio statistic, derived from the likelihood of a model with G classes divided by the likelihood of a model with $G+1$ classes, is only asymptotically χ^2-distributed if all possible response patterns have a reasonable chance of appearing.

The advantage of the LCA over the RM is that different sets of item parameters are allowed in different classes. A disadvantage of the LCA is that it does not allow any variation between the persons´ parameters (abilities) within a latent class. This assumption of constant response probabilities for all individuals in a latent class has proven too restrictive for many purposes.

1.3 The Mixed Rasch Model (MRM)

The MRM "combines the theoretical strength of the Rasch model with the heuristic power of latent class analysis. It assumes that the Rasch model holds for all persons within a latent class, but it allows for different sets of item parameters between the latent classes" (Rost, 1990, p. 271). Thus, the MRM is the supermodel of both the LCA and the RM. The response function of the MRM is described by the following equation:

$$p(xvi) = \sum_{g}^{G} \pi_g \frac{e^{(\xi_{vg} + \alpha_{ig})}}{1 + e^{(\xi_{vg} + \alpha_{ig})}} \qquad (6)$$

The response probability $\pi_{i|g}$ from Equation (4) is now rewritten in accordance with the RM. It is obvious that in the case of only one latent class the MRM is reduced to the ordinary RM. In the case of different classes but without variation of the ability parameters within each class, the MRM correspondingly becomes a simple LCA model.

The parameters of the MRM can be estimated by means of an extended EM-algorithm with conditional maximum likelihood estimations of the item parameters in the M-step (see Rost, 1990, 1994). To assess the model fit of a given solution with G latent classes, again the AIC and BIC-index are computed.

2. Identification of Guessing Behavior on the Basis of Item Respose Theory (IRT)

Achievement tests standardly consist of multiple choice items, each having J choices. A person´s probability of solving such items by guessing is $p=1/J$. If there is any subsample of persons who have guessed in a given data set, the assumption of the ordinary RM that all item difficulties are constant for all persons is no longer valid. Those guessing examinees´ item parameters will be different from those of the

remaining persons who use cognitive skills to solve the items. In correspondence with this deviant behavior of a subsample a significance test (for instance Andersen´s likelihood-ratio test) will indicate that the RM does not hold for the whole data set. If the guessing examinees are excluded, the RM will fit the residual sample´s data.

A common strategy for indirectly infering guessing behavior from given response vectors of N subjects has two steps. In the first step it is assumed that the RM is valid for all persons, and the model parameters are then estimated. On the basis of these estimates an ICC is plotted for each item. In the second step the proportion of correct answers for each score (ability) group is plotted against the ICC of a difficult item. "Guessing behavior is assumed to be operating when test performance for the low performing score groups exceeds zero". (Hambleton & Swaminathan, 1989, p.161). Figure 2 depicts such a graph, where the proportions of correct answers (indicated by dots) are plotted against the ICC, which indicates the response probability under the assumption that the RM is valid.

As a consequence of these deviations between empirical and expected success probabilities, educational researchers and psychometricians often apply the three-parameter logistic model (Birnbaum, 1968) whereby the probabilty of person v to solve item i ist described by:

$$p(x_{vi}) = \gamma_i + (1 - \gamma_i)\frac{e^{\beta_i(\xi_v + \alpha_i)}}{1 + e^{\beta_i(\xi_v + \alpha_i)}}. \qquad (7)$$

β_i is the so-called discrimination parameter of item i. γ_i, which is more important for our current purpose, is the so-called guessing parameter.[2] The introduction of a guessing parameter γ_i consequentially means that the lower asymptote of the ICC is in general greater than zero. Thus, the ICC in Figure 2 would converge with decreasing abilities against γ_i but not zero. Actually, γ_i is normally less than the real guessing probability which in the case of a multiple choice item with J choices is $p=1/J$. This phenomenon is explained by Lord (1974), who argued that some examinees with low abilities do not guess in the case of a difficult item but choose the most attractive distractor of the item.

[2] β_i describes variations in the discrimination power of different items. In the ordinary RM all ICCs are nonintersecting curves that differ only by a translation along the the latent continuum. These items vary only in their difficulty. The two-parameter model additionally allows a variation among the slopes of the ICCs, that is a variation among the discrimination powers of different items. The higher β_i the better the discrimination power of item i.

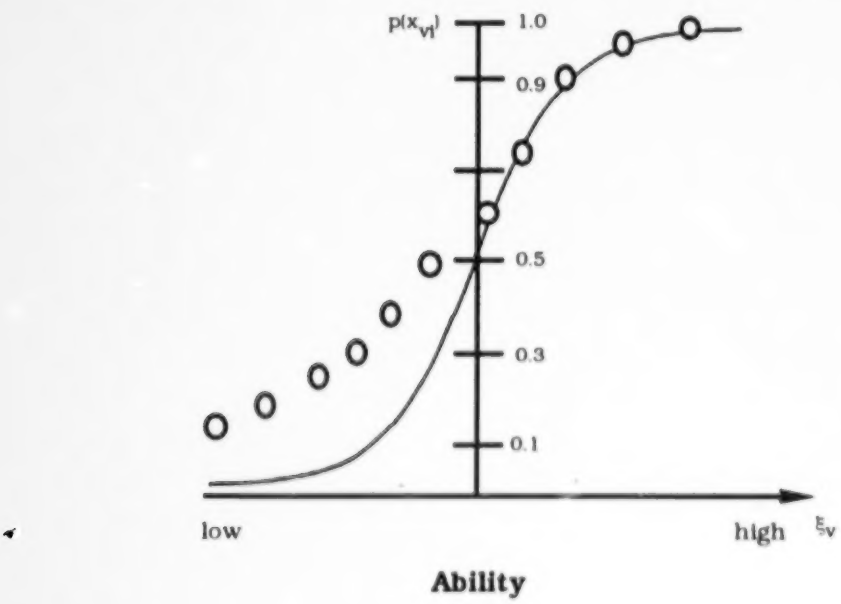

Figure 2
Plot of expected solving probabilities (represented by the ICC) and empirical proportions of correct answers (indicated by dots) for an item which provoked guessing behavior in low ability groups

Problems or disadvantages of the three-parameter model as pointed out by Kubinger (1988) are:

(1) The assumption of specific objectivity is abandoned, i.e. estimations of individual parameters are no longer independent of the given subset of items.

(2) The parameters estimated by means of the unconditional maximum likelihood method are not consistent. Optimal and stable estimates are only possible if the sample sizes reached more than 1000 examinees and more than 50 items.

(3) Comparisons between the ordinary RM and the two- or three-parameter model are difficult. Only in the case of more than 30 items is a likelihood-ratio test similar to Andersen's test applicable.

Contrary to the two- and three-parameter logistic model, the MRM includes, at least within the latent classes, all features of the ordinary RM, for example specific objectivity and stable parameter estimations by means of conditional maximum likelihood method, even if the number of items is small. The AIC and BIC-Index give an opportunity to compare different models with varying numbers of classes.

2.1 Guessing Behavior and MRM - a Simulation Study

In the following simulation, we assume a given empirical data set containing the responses of N examinees on m items with J choices. Some of the examinees (a subsample of n ($n<N$) individuals) have applied a guessing strategy to all m items. Analyzing these data by means of the MRM we would expect two latent classes, one of which could be characterized as the "guessing class," with item expectation values equal to the random probability ($p=1/J$) and low or, for an indefinite number of persons, no variation between the item parameters. Another possible characteristic for this class is the expectation that, for an indefinite number of items, all individuals should have the same ability, i.e. no variance between the person parameters should occur. This means that within the "guessing class" the RM is reduced to a LCA model with only one class. Dealing with finite samples of m items and N persons we would normally find a small variation of person and item parameters in empirical data, however, caused by random differences not directly related to ability.

The item and individual parameters of the second class should vary significantly, and the expectation values should deviate substantially from the random probability. For this group we would assume that we have measured the intended trait, e.g. scholastic achievement. To illustrate these assumptions and expectations we will analyze a simulated data set below. Let us start with an intelligence test consisting of 24 items with 5 choices. Here, the random probability of success is p=0.20. Suppose the existence of two latent classes, one of which contains guessing examinees, the other subjects who use their cognitive abilities to solve the items. The response probabilities (item means) of the 24 items in this latter class are assumed to be overall p=0.7 for the first 8 items, p=0.5 for the second 8 items and p=0.3 for the last 8 items. According to different "subjects´ capabilities" these probabilities of success should vary between different ability groups, as shown in detail in Table 1.

All members of group 1 have the highest probabilities of success, followed by group 2 and so on up to group 5, i.e. the simulated abilities decrease from groups 1 to 5. In group 6 we assume the guessing strategy and fix all response probabilities at p=0.2, which is the random probability of solving a multiple choice item with five choices.

The corresponding data for each item were first generated for 600 guessing examinees by means of a small simple BASIC program. Whether a subject of this class received a 0 (item not solved) or a 1 (item solved) was decided by drawing random numbers from

the standard normal distribution. According to the value drawn the density of the standard normal distribution was calculated from infinite to the current value. If this density was less than or equal to 0.8, the person´s response on the item was scored 0, otherwise it was scored 1.

Table 1
Response probabilities and sample sizes of the 6 simulated groups. Group 6 consists of "guessing examinees."

Group (v)	$p(x_1=1 \mid \xi_v)^a$	$p(x_2=1 \mid \xi_v)^b$	$p(x_3=1 \mid \xi_v)^c$	N^d
1	0.9	0.7	0.5	540
2	0.8	0.6	0.4	540
3	0.7	0.5	0.3	540
4	0.6	0.4	0.2	540
5	0.5	0.3	0.1	540
6	0.2	0.2	0.2	600

[a]Response probability of a subject in group v on an item with an overall solving probability of p=0.7 in the "non-guessing group"
[b]Response probability of a subject in group v on an item with an overall solving probability of p=0.5 in the "non-guessing group"
[c]Response probability of a subject in group v on an item with an overall solving probability of p=0.3 in the "non-guessing group"
[d]Number of persons per group

The remaining simulated persons from class 1 up to class 5 were generated with the same procedure, but, depending on their supposed solving probabilities, the criterion for whether they received a 0 or a 1 was varied. For instance members of group 1 (with "high abilities") were scored 0 for the first 8 items, if the density between infinite and the value drawn was less than or equal to 0.1; otherwise they received a 1. As another example, members of group 5 (with "low abilities") were scored 0 for the last 8 items, if the density between infinite and the value drawn was less than or equal to 0.9; otherwise they received a 1.

This procedure was applied to all groups, resulting in a total of 3300 simulated subjects, 600 of them "guessing examinees". Their responses on the 24 simulated items were analyzed with the PC program MIRA by Rost and Davier (1992). Table 2 contains goodness-of-fit statistics for different solutions from one up to four classes.

Table 2
Goodness-of-fit statistics for different solutions of the MIRA-analysis

G[a]	log L[b]	k[c]	BIC[d]
1	-48594.98	47	97570.70
2	-48270.50	93	97294.38
3	-48221.83	139	97569.68
4	48185.32	185	97869.29

[a]number of latent classes; [b]log-likelihood; [c]number of estimated independent parameters; [d]Best Information Criterion (Bozdogan, 1987)

According to the BIC-index the two-class solution displays the best fit. The first class consists of 627 members, the second of 2673. The assignment of an examinee to the different classes was executed with respect to his or hers response pattern; any subject was assigned to that latent class where, under the condition of his or hers response vector, the membership probability was highest. 6.4% of all persons in total were misclassified, that is, were assigned to the guessing class although they had been simulated as non-guessers, or were assigned to the non-guessing class despite being simulated as guessers.[3]

The interpretation of the two latent classes can be drawn from the graphical representation of the item expectation values in Figure 3. Congruent with the simulation suggestions, class 1 is characterized by item expectation values near the random probability of p=0.20. The corresponding item parameter estimations vary between α_{ig}=-0.22 and α_{ig}=+0.26[4], the variance of the individual parameters amounts to $V(\xi_g)$=0.36.

The second latent class also shows item expectation values in accordance with the simulated assumptions. The expectation values for the first 8 items are near p=0.70, for the second 8 items near p=0.50 and for the last 8 items near p=0.30. The corresponding item parameters vary between α_{ig}=-1.02 and α_{ig}=+1.02. The variance of

[3]These misclassifications are caused by random processes; a guessing person, for instance, can by chance reach a response vector, which will cause him or her to be assigned to the class of non-guessing examinees.
[4]In MIRA the sum of all item parameters is standardized to 0. In the case of an infinite number of examinees all item paramters should be equal to 0 because of the above norming condition.

the person parameters is $V(\xi_g)$=0.64, which is significantly higher (F(2699,599)=1.78, p<0.001) than the variance in the guessing class.

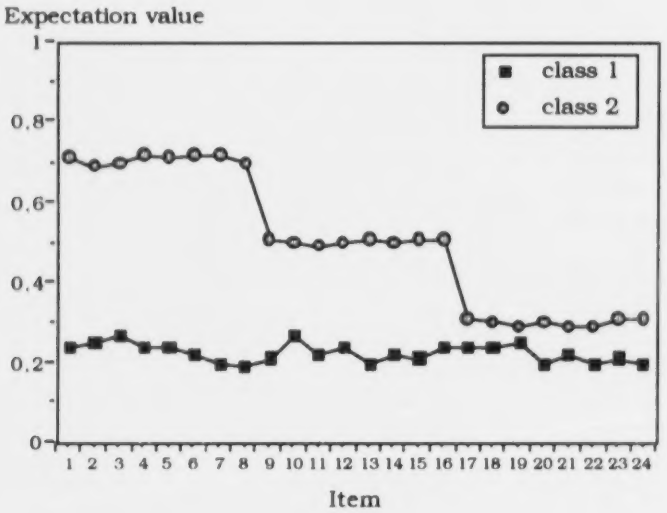

Figure 3
Expectation values of the 24 simulated items within the two latent classes

In summary, we have illustrated by means of a simulation study that the MRM is a potential tool to identify examinees who apply a guessing strategy to a set of multiple choice items.

3. Identification of Guessing Behavior - an Empirical Study

In our study 5641 7th grade students were tested with different scholastic achievement and intelligence tests, all of which were speed tests.[5] As an example we analyzed the results in a biology test consisting of 23 multiple choice items, each having 5 choices. Right answers were scored with 1, wrong anwers with 0. The investigated sample consists of N=2889 students (50.7% females) from a state of the former Federal Republic of Germany (FRG), and N=2752 students (53.4% females)

[5]This current investigation is part of a longitudinal study called "Educational Processes and Psycho-Social Development in Adolescence (BIJU)" which began in 1991. The following institutions are involved: Institute for Science Education (IPN), Kiel; Max Planck Institute for Human Development (MPI), Berlin; Humboldt University, Berlin and Martin Luther University, Halle. The project leaders are Prof. Dr. J. Baumert (IPN) and Prof. Dr. P.M. Roeder (MPI).

from two states of the former German Democratic Republic (GDR). In accordance with our assumptions, we expected at least two latent classes:

Class 1: In this class, the 23 items should test the factor "biology knowledge". The item expectation values should deviate significantly from the random probability, and the amounts of variance between item parameters and between individual parameters should be significantly higher than those in class 2.

Class 2: In this class we expected to find all those examinees who applied a guessing strategy to solve the tasks. Most of their item expectation values should be equal to the random probability p=0.20, when five choices are present.[6] The amount of variance of the item and individual parameters should be very small.

Table 3 contains goodness-of-fit statistics for different MIRA-solutions from one to four classes.

Table 3
Goodness-of-fit statistics for different solutions of the MIRA-analysis for the biology test

G[a]	log L[b]	k[c]	BIC[d]
1	-75545.51	45	151479.64
2	-75100.37	89	150969.36
3	-74799.16	133	150747.02
4	-74648.18	177	150817.47

[a]number of latent classes; [b]log-likelihood; [c]number of estimated independent parameters; [d]Best Information Criterion (Bozdogan, 1987)

According to the BIC-index, the three-class solution fits the data better than all other solutions. Figure 4 shows the item expectation values for the three latent classes.

Class 1 (26.1% of the whole sample) is characterized by the fact that most of the item expectation values are similar to the random probability p=0.20. Only very easy items, (items 4, 6, 7, 8, 9 and 18), with expectation values greater than p=0.70 in the other

[6]Only very easy items with obvious solutions should form an exception. For these items we assume expectation values greater than the random probability.

classes, have values substantially higher than 0.20. This first class might have used a guessing strategy to solve the items. Only easy tasks provoked a temporary change in this strategy. The item parameters vary between α_{ig}=-0.86 and α_{ig}=+1.94, the variance of the individual parameters is $V(\xi_g)$=0.48. The range of the item parameters is greater than that of the simulated data, which can be explained by the fact that particularly easy items in the empirical data did not provoke a guessing strategy. The mean of solved items is M=7.75.

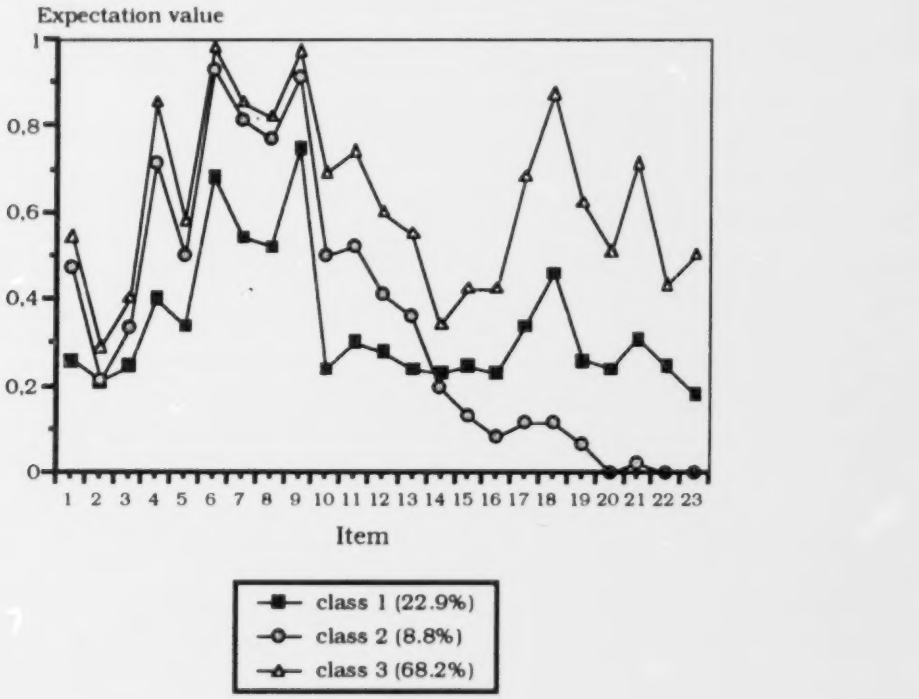

Figure 4
Item expectation values for the three-class solution from MIRA

Class 2 (64.9% of the whole sample) shows expectation values, all deviating from the random probability (p=0.20). These results support the hypothesis that, in this latent class, we have actually tested the factor "biology knowledge" and not guessing behavior. The item parameters vary between α_{ig}=-1.45 and α_{ig}=+3.04, the variance of the individual parameters is $V(\xi_g)$=0.83. The mean of solved items is M=14.14 and therefore significantly higher than in class 1 (t=70.70, df=5126, p=0.000).

Class 3 (8.9% of the whole sample) differs from class 2 both quantitatively and qualitatively. "Quantitative" means that these examinees show lower item expectation values for the first 14 items than those in class 2. This can be explained by a lower level of knowledge in class 3. "Qualitative" means that, beginning with item 15, a second trait, namely the processing speed, has influenced the item response probabilities. The test was administered as a speed test, and obviously the persons in class 3 were not able to solve the last items because of their slow processing speed. The item parameters in this class vary between α_{ig}=-6.91 and α_{ig}=+4.29 and the variance of the individual parameters amounts to $V(\xi_g)$=0.62. The mean of solved items is M=8.36 which corresponds approximately to the mean in class 1 and is significantly lower than in class 2 (t=40.32, df=4442, p=0.000).

In summary, the results of the MIRA-analysis support the hypothesis that a latent guessing class exists. This assumption is confirmed (1) by the profile of the item expectation values within this class and (2) by the reduced variance within it. Pairwise F-tests revealed that the variance within the guessing class was significantly smaller than within the other classes (p<0.001).

3.1 The Relationship between "Guessing Behavior" and Cognitive and Motivational Variables

The goal of the following step was to analyze the relationship between guessing behavior and cognitive, motivational and self-related variables. We did not analyze differences in cognitive variables among the three latent classes by means of an intelligence test, because several MIRA-analyses of subscales from the *Intelligence Structure Test* (in German: *Intelligenzstrukturtest IST*, Amthauer, 1953) and from the *Cognitive Ability Test* (in German: *Kognitiver Faehigkeitstest KFT 4-13*, Heller, Gaedike & Weinlaeder, 1976) revealed that, again, different strategies were used to solve the items. Instead of an intelligence test, we analyzed differences in scholastic achievement measured by grades.

Table 4 contains the results of our mean comparisons among the three latent classes with respect to the biology grade and the sum of grades in mathematics and German. These variables were both standardized beforehand, so that values below zero indicate a low level of scholastic achievement (below the mean), and values above zero stand for a high level of achievement (above the mean). As expected, the guessing class

shows the lowest achievement level both in biology and in the combination of mathematics and German. The highest achievement is shown by class 2, in which we measured the desired variable "biology knowledge", followed by class 3, which was characterized by a slow processing speed.

Table 4
Class means of different achievement variables and motivational factors

	class 1	class 2	class 3
Math and German grades	-0.66	0.20	-0.32
Biology grade	-0.61	0.19	-0.34
Biology-specific anxiety	0.37	-0.12	0.06
Self-concept of ability in biology	-0.39	0.13	-0.10

Oneway analyses of variance revealed that all variables differ significantly among the three groups ($p<0.01$). Using Tukey's hsd-test we obtained the result, that all pairwise comparisons were significant ($p<0.01$) for all variables.

Table 4 also provides information regarding the standardized means of the three latent classes with respect to the biology-specific anxiety and the self-concept of ability in biology. Anxiety was measured by means of a test by Helmke (1992) containing 23 items with 5-point ratings for each item. The domain-specific self-concept of ability was tested by means of a short scale by Jerusalem (1984) consisting of 5 items with 4-point ratings for each item. The reliability of both scales was satisfactory (Cronbach´s α=.93 for the anxiety scale and α=.87 for the self-concept scale). Table 5 presents some item examples. These two affective variables were chosen to obtain information regarding whether the guessing students show characteristics which, in addition to a lower cognitive level, inhibite the learning and performing process in school.

According to the self-concept of ability a meta-analysis by Hansford and Hattie (1982) shows a correlation coefficient of r=.42 between self-concept and scholastic achievement, which indicates a strong relationship between these two variables. An appropriate interpretation of this relationship is that a higher self-concept corresponds to a higher level of aspiration, which stimulates persistence in the students´ learning processes and leads to higher knowledge and performance on achievement tests.

With respect to the relationship between anxiety and achievement, the conclusions of the relevant literature can be interpreted as indicating that a high level of anxiety does not have any strong positive or negative influence on the learning process but on the concrete performance situation, in which students have to solve tasks (Schnabel & Gruehn, 1993).

Based on this theoretical background, the results summarized in Table 4 are in accordance with our expectations. Compared with the two other latent classes the guessing examinees expressed a higher level of anxiety and a lower self-concept of ability. It is plausible that the higher level of anxiety in the performance situation (solving the 23 items of the biology test), combined with lower cognitive capabilities, provoked the guessing behavior. The random strategy might have been chosen to cope with the items, which were too difficult.

Table 5
Item examples for the scales *self-concept of ability in biology* (Jerusalem, 1984) and *biology-specific anxiety* (Helmke, 1992)

Self-concept of ability in biology
1. I would like biology if this subject were not so difficult.
2. Even if I do my best in biology, I do not perform as well as the other students in my class.
(ratings from 1="do not agree" up to 4="strongly agree")
Biology-specific anxiety
Please remember the last class test in biology. How did you feel?
1. I doubted my abilities.
2. I imagined who among the other students would perform worse than I.
(ratings 1="do not agree" up to 4="strongly agree")

4. Summary and Discussion

Psychological hypotheses addressing different cognitive strategies often seem to be incompatible with psychometric models concerning test behavior. A common test model like the ordinary Rasch model, which assumes that a set of items measures the

same trait, is not applicable to the case where different cognitive strategies are used to solve the items. In the current study we introduced the MRM as a statistical possibility for detecting such different processing strategies in achievement tests. Applying the MRM to simulated and empirical data we tested and confirmed the hypothesis that at least two strategies can be applied to scholastic achievement tests consisting of multiple-choice items: a guessing strategy and a strategy based on knowledge. Further analyses provided information that guessing behavior is shown by students with lower-level cognitive abilities and a higher level of anxiety, who perhaps use the "random strategy" to cope with the items too difficult for them.

In addition to this class with examinees who guess, we identified two other classes: one of which could be described simply as those persons whose responses were based on knowledge, the other of which could be characterized by the fact that their item response probabilities were influenced by two dimensions, namely biology knowledge and processing speed. This result further demonstrates that the MRM can also provide researchers with information about different processing strategies in different groups.

As a consequence of our approach, the interpretation of students´ scores on achievement tests should include two steps. In the first step researchers have to identify the applied strategy of a given examinee and in the second step calculate the individual (ability) parameter for the identified latent variable. The MRM allows such a qualitative and quantitative analysis of given response vectors.

References

Birnbaum, A (1968). Some latent trait models and their use in infering an examinee´s ability. In F.M. Lord & M.R. Novik (Eds.). *Statistical theories of mental test scores, 395-479*. Reading/Mass.: Addison-Wesley.

Bozdogan, H. (1987). Model selection and Akaike´s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.

Hambleton, R.K. & Swaminathan, H. (1989, 4th Ed.). *Item Response Theory. Principles and applications*. Boston: Kluwer Nijhoff Publishing.

Hansford, B.C. & Hattie, J.A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 53*, 75-91.

Heller, K., Gaedike, A.-K. & Weinlaeder, H. (1976). *Kognitiver Faehigkeitstest KFT 4-13*. Weinheim: Beltz.

Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Goettingen: Hogrefe.

Jerusalem, M. (1984). *Selbstbezogene Kognitionen in schulischen Bezugsgruppen - eine Laengsschnittstudie*. Departement of Psychology, Free University, Berlin.

Kirk, R. E. (1968). *Experimental Design: Procedures for the behavioral science*. Belmont, Californien: Brooks/Cole.

Koeller, O., Rost, J. & Koeller, M. (1994). Individual differences in solving "cube tasks" of the IST and IST-70. *Zeitschrift fuer Psychologie, 202*, 63-82.

Kubinger, K.D. (1988). Aktueller Stand und kritische Wuerdigung der Probabilistischen Testtheorie. In K.D. Kubinger (Ed.). *Moderne Testtheorie*, 19-83. Weinheim, Basel: Beltz.

Lazarsfeld, P.F. (1950). Logical and mathematical foundations of latent structure analysis. In S.A. Stouffer, L. Guttman, E.A. Suchan, P.F. Lazarsfeld, S.A. Star & J.A. Clausen. *Studies in social psychology in world war II*. Vol. IV. Princeton/N.J: Princeton Univ. Press.

Lazarsfeld, P.F. & Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Miflin Co.

Lord, F.M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39*, 247-264.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Nielson & Lydiche.

Rost, J. (1990). Rasch models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

Rost, J. & Davier, M.v. (1992). MIRA - *A PC program for the mixed Rasch model. - User Manual*. Kiel: Institute for Science Education (IPN) at the University of Kiel.

Rost, J. & Davier, M. v. (1993). *Das Mixed-Rasch-Modell und die Identifikation von unskalierbaren Personen*. Paper presented at the 1. meeting of the methodology section within the "Deutsche Gesellschaft fuer Psychologie". Kiel, 27.-29.09.1993.

Rost, J. & Georg, W. (1991). Alternative Skalierungsmoeglichkeiten zur klassischen Testtheorie am Beispiel der Skala "Jugendzentrismus". *Zentralarchiv fuer Empirische Sozialforschung, Informationen, 28*, 52-75.

Rost, J. (1994, i. p.). Mixture distribution models. In W. van der Linden & R. Hambleton (Ed.). *Handbook of modern IRT*. New York: Springer.

Schnabel, K. & Gruehn, S. (1993). *Fachspezifische Leistungsangst und ihr Einfluss auf die Leistungsentwicklung*. Paper presented at the 49. meeting of the "Arbeitsgruppe fuer empirisch-paedagogische Forschung (AEPF)." Vienna, 13.-15.09.1993.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.